Behavioral/Cognitive

The Social Brain Automatically Predicts Others' Future Mental States

Mark A. Thornton,[1,2] Miriam E. Weaverdyck,[1] and Diana I. Tamir[1,2]
[1]Department of Psychology and [2]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08540

Social life requires people to predict the future: people must anticipate others' thoughts, feelings, and actions to interact with them successfully. The theory of predictive coding suggests that the social brain may meet this need by automatically predicting others' social futures. If so, when representing others' current mental state, the brain should already start representing their future states. To test this hypothesis, we used fMRI to measure female and male human participants' neural representations of mental states. Representational similarity analysis revealed that neural patterns associated with mental states currently under consideration resembled patterns of likely future states more so than patterns of unlikely future states. This effect manifested in activity across the social brain network and in medial prefrontal cortex in particular. Repetition suppression analysis also supported the social predictive coding hypothesis: considering mental states presented in predictable sequences reduced activity in the precuneus relative to unpredictable sequences. In addition to demonstrating that the brain makes automatic predictions of others' social futures, the results also demonstrate that the brain leverages a 3D representational space to make these predictions. Proximity between mental states on the psychological dimensions of rationality, social impact, and valence explained much of the association between state-specific neural pattern similarity and state transition likelihood. Together, these findings suggest that the way the brain represents the social present gives people an automatic glimpse of the social future.

Key words: emotion; functional magnetic resonance imaging; predictive coding; repetition suppression; representational similarity analysis; social cognition

Significance Statement

When you see a ball in flight, your brain calculates, not just its static visual features such as size and shape, but also predicts its future trajectory. Here, we investigated whether the same might hold true in the social world: when we see someone flying into a rage, does our brain automatically predict their social trajectory? In this study, we scanned participants' brain activity while they judged others' mental states. We found that neural activity associated with a given state resembled activity associated with likely future states. Additionally, unpredictable sequences of states evoked more brain activity than predictable sequences, consistent with monitoring for, and updating from, prediction errors. These results suggest that the social brain automatically predicts others' future mental states.

Introduction

Social life requires people to predict the future: we must predict what others think, feel, and do to interact with them successfully. Much of how we behave toward other people depends on what we anticipate they will do next. We intuitively know to tread gingerly around frustrated colleagues, to anticipate and allay children's fears, or to order-in comfort food after a friend's stressful job interview. Humans have many means for perceiving others' current thoughts and feelings: we read facial expressions, hear tone of voice, and interpret situational contexts (Zaki et al., 2009; Barrett et al., 2011; Becchio et al., 2018). However, we cannot see or hear others' emotions before they happen. How do we satisfy our need to glimpse the social future?

We propose that the brain meets this need by automatically predicting the social future when considering the social present. Our perceptual system engages in this kind of reflexive prediction (Rao and Ballard, 1999; Hohwy et al., 2008; Vuust et al., 2009):

Received June 6, 2018; revised Oct. 17, 2018; accepted Oct. 23, 2018.

Author contributions: M.A.T. wrote the first draft of the paper; M.A.T., M.E.W., and D.I.T. edited the paper; M.A.T., M.E.W., and D.I.T. designed research; M.E.W. performed research; M.A.T. and M.E.W. analyzed data; M.A.T. and D.I.T. wrote the paper.

This work was supported by the National Institutes of Mental Health–National Institutes of Health (Grant R01MH114904 to D.I.T.). We thank Aaron Kurosu, Betsy Levy Paluck, Judith Mildner, Aidan O'Donnell, Sarah Pan, and Zidong Zhao for advice and assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Mark A. Thornton, Department of Psychology, Peretsman Scully Hall, Princeton University, Princeton, NJ 08540. E-mail: mthornto@princeton.edu.

https://doi.org/10.1523/JNEUROSCI.1431-18.2018

 0270-6474/19/390140-09$15.00/0

when we observe a ball in flight, we not only represent its current location, but also automatically predict its trajectory. The brain compares its predictions with incoming sensory information, calculates mismatch between the two, and adjusts subsequent predictions accordingly. This algorithm, which is called predictive coding, offers a biologically plausible mechanism for implementing Bayesian-optimal prediction (von Helmholtz, 1910; Gregory, 1980; Friston and Kiebel, 2009; Clark, 2013). Researchers have already begun to demonstrate that the brain may employ an algorithm like predictive coding in the social and affective domains (Koster-Hale and Saxe, 2013; Barrett, 2017; Theriault et al., 2017). For example, research on the human mirror neuron network suggests that predictive coding of others' actions may help us infer their intentions (Kilner et al., 2007).

In this study, we focused on how people use knowledge about a person's current mental state to predict their future mental states. Experience-sampling data indicate that a person's current mental state reveals much about their likely future states (Thornton and Tamir, 2017). For example, if a person currently feels rage, then they are more likely to next feel disgust than gratitude. Moreover, perceivers successfully make use of this regularity: people can make accurate explicit predictions up to two emotions into others' futures.

People may predict others' future states with such high fidelity because prediction is built into the way that people represent social knowledge. That is, these social predictions occur automatically because the brain directly incorporates representations of likely future states into representations of others' current states. If so, then whenever someone thinks about a mental state, their brain's representation of that state would include, not only its static features, but also features of predicted future states. In other words, if people use predictive coding to anticipate others' states, then when one observes a friend fly into a rage, the brain should automatically predict that person's emotional trajectory. Here, we used representational similarity analysis (Kriegeskorte et al., 2008) and repetition suppression in fMRI data to test this hypothesis.

In addition to testing whether the brain reflexively predicts others' mental states, we also tested a theoretical model developed to explain specifically how it might make these predictions (Tamir and Thornton, 2018). Specifically, we tested whether a dimensional model of mental state representation serves as scaffolding for social prediction. In this model, people represent mental states using three psychological dimensions (Tamir et al., 2016): rationality, social impact, and valence. Each mental state is defined by its coordinates on these dimensions (e.g., envy is emotional, socially arousing, and negative). This organization should facilitate automatic social prediction because a state's position on these dimensions determines its transitional probabilities (Thornton and Tamir, 2017). That is, the closer two states are on each dimension, the greater the likelihood of transitions between them. Here, we test the extent to which proximity on these dimensions statistically mediates the neural predictions of future states from current states.

Materials and Methods

Code accessibility

Data and code have been deposited on the Open Science Framework (https://osf.io/x32te/) and are freely available. We report how we determined sample size, all data exclusions, all manipulations, and all measures in the study. We drew on previously published studies for two datasets: the scores of each mental state on four psychological dimensions (Tamir et al., 2016) and ratings of transitional probabilities between mental states (Thornton and Tamir, 2017). Data were analyzed with a combination of common fMRI software packages detailed below and custom in-house code in MATLAB (RRID: SCR_001622) and R (RRID:SCR_001905).

Participants

Imaging participants ($N = 29$) were recruited from Princeton University Credit and Paid Study Pools. One participant was excluded due to data loss in the image reconstruction process. The remaining participants (17 female, 11 male; age range 18–22, mean age = 19.6) were right-handed, neurologically normal, fluent in English, and had normal or corrected-to-normal vision. Participants provided informed consent in a manner approved by the Princeton University Institutional Review Board.

Sample size was determined via resampling-based power analysis using data from a previous study of similar design and content (Tamir et al., 2016). In that study, participants made judgments on ~60 mental states (including the 15 studied here). Three orthogonal psychological dimensions were found to explain similarity between the activity patterns elicited in this task: rationality, social impact, and valence. Here, we targeted the smallest significance effect size ($r = 0.12$, for valence) and drew bootstrapped samples of participants from the original study to determine statistical power for replicating this effect with varying sample sizes. Due to the subset of states chosen, the expected effect size was larger than in the earlier dataset, yielding a target of 28 participants to achieve 95% power.

Behavioral participants ($N = 29$) were recruited online via Amazon Mechanical Turk. These participants rated the pairwise perceived similarity between the mental states presented in the imaging study. One participant was excluded for indicating that English was not their native language and that their fluency was less than excellent (leaving $N = 28$; 10 female, 18 male; age range 20–64, mean age = 34.3). Sample size for this group was chosen to match the high reliability of ratings of transitional probabilities collected in a previous study (Thornton and Tamir, 2017) that were reused as an independent variable here.

Stimuli

The stimuli in the imaging paradigm consisted of 15 different mental state terms: consciousness, desire, disgust, distrust, drunkenness, embarrassment, exhaustion, friendliness, lust, patience, playfulness, satisfaction, sleepiness, trance, and transcendence. These states were selected from a larger set of 60 used in a previous study (Tamir et al., 2016). We selected these particular states to maximize asymmetries in transitional probabilities (Thornton and Tamir, 2017) and thereby maximize anticipatory repetition suppression effects and asymmetries in neural pattern similarity. Each state was paired with 30 brief scenarios that had been pretested to elicit the state in question (e.g., happiness: "pet a puppy"). These were selected from larger sets of 36 scenarios for each state via genetic algorithm. This ensured that the extent to which the scenarios elicited their given states was as high as possible, but also balanced across states (Tamir et al., 2016). Additionally, the algorithm sought to minimize differences in variance in the scenario appropriateness across states and variance in the average character length of the scenarios across states.

Experimental design and statistical analysis

Experimental design. In the imaging paradigm, participants rated how much a given scenario would elicit a particular state in another person. The target of mentalizing was deliberately generic, describe as only "another person" or "a person" in the instructions. On each trial, a participant might be presented with the state "desire" and the scenario "walking into a candy store." The state initially appeared alone for 250 ms and then the scenario and rating scale appeared below it and participants had 2.5 s to read and respond. Participants made their ratings on a 1-to-5 Likert-type scale using a button box positioned in their left hands. This was followed by a 250 ms fixation period before the start of the next trial. There were 225 trials in each of four runs, for a total of 900 trials. In each run, each mental state was presented 15 times and proceeded by every other mental state exactly once. An additional 6 s fixation period was allowed at the end of each run to ensure capture of hemodynamic responses from the final trials. The order of presentation was counterbalanced and optimized for continuous-carryover repetition suppression

(Aguirre, 2007) with respect to previously rated transitional probabilities (Thornton and Tamir, 2017) via de Bruijn cycles (Aguirre et al., 2011). Each scenario was presented twice over the course of the experiment, either in the two even or in the two odd runs. Outside of the scanner, participants rated their perceptions regarding the transitional probabilities between the 15 mental states in the imaging paradigm, as well as how long they thought each state typically lasted.

After the scanning session, participants completed measures of the Autism Spectrum Quotient (Baron-Cohen et al., 2001b), the UCLA Loneliness scale (Russell, 1996), a measure of social network size, a single-item extraversion measure, the Narcissistic Personality Inventory (Raskin and Hall, 1979), the MOS social support survey (Sherbourne and Stewart, 1991), and the Reading the Mind in the Eyes task (Baron-Cohen et al., 2001a), as well as demographics and open-ended feedback about the experiment and its purpose. These measures were collected for cross-study analyses and were not analyzed as part of the current investigation.

Ratings of the perceived similarity between mental states were provided online by a separate sample of participants. These participants were recruited using TurkPrime (Litman et al., 2017) and then directed to a Qualtrics-based survey. A "captcha" image was used to help protect the survey from automated responding. In the survey, participants rated the perceived similarity between each pair of mental states within the set of 15 states presented in the imaging study. Ratings were made using a continuous line scale anchored at "not at all similar" and "very similar." To encourage the possibility of asymmetric similarity ratings, the prompt was phrased "how similar is (state 1) to (state 2)?" Additionally, to minimize explicit attempts to be consistent across asymmetries, ratings corresponding to the upper and lower triangular portions of the similarity matrix were separated into separate blocks. This procedure helped to ensure that trials featuring the same mental states would rarely be presented close together. The order of the blocks and the trials within each block were independently randomized for each participant. Participants reported their demographics (age, gender, race, ethnicity, and English-language proficiency) and had the option to provide open-ended feedback at the end of the study.

Imaging procedure. Imaging data were acquired at the Princeton Neuroscience Institute using a 3 tesla Siemens Skyra scanner with a 64-channel head coil. Functional gradient-echo echoplanar images were obtained from the whole brain using a simultaneous multislice imaging procedure and online motion correction (66 interleaved slices of 2 mm thickness, TR = 1500 ms, TE = 32 ms, flip angle = 70°, in-plane resolution = 2.00 × 2.00 mm, matrix size = 192 × 192 voxels, 162 measurements per run). Functional images were preprocessed using a multipackage imaging pipeline: corrections for slice timing and head motion were performed using FSL (RRID:SCR_002823) (Jenkinson et al., 2012), normalization to the ICBM 152 template (RRID:SCR_008796) was completed using SPM8's DARTEL (Ashburner, 2007), and smoothing and the linear modeling were performed using SPM12 (Wellcome Department of Cognitive Neurology, London, UK) as part of the SPM12w package (https://github.com/wagner-lab/spm12w) (SPM, RRID:SCR_007037).

The general linear model (GLM) was used to prepare each participant's data for representational similarity analysis. Each mental state was modeled as a condition of interest (15 total) using a boxcar regressor that began on each trial when the name of the state appeared and lasted until the participant responded or the response window ended, whichever came first. The regressors were convolved with a canonical hemodynamic response function (HRF) and entered into the GLM along with additional nuisance covariates: run means and linear trends and six motion realignment parameters.

Behavioral data analysis. Behavioral data were analyzed to assess the quality of participants' engagement with the imaging task. This included calculation of descriptive statistics for response rate, average response, average standard division of responses with respect to mental state, and average reaction time. Additionally, because participants responded to each scenario twice across the course of the experiment, we could calculate a measure of test–retest reliability by correlating their first and second rating of the same stimulus. The correlation values were *r*-to-*z* transformed and entered into a one-sample *t* test versus zero to assess the statistical significance of the reliability at the group level.

Representational similarity analysis. Neuroimaging data were subjected to representational similarity analysis (Kriegeskorte et al., 2008) to test whether transitional probabilities between states could explain the similarity between corresponding patterns of brain activity. That is, if a pattern of brain activity encodes predictions about likely future states, then the pattern associated with each state should resemble the patterns associated with likely future states more than the patterns associated with unlikely future states. This hypothesis was tested at two levels of analysis: in a whole-brain similarity searchlight procedure (Kriegeskorte et al., 2006) and across the entire social brain network.

In the searchlight analysis, a small, approximately spherical region with a 4-voxel radius was centered at each voxel in the brain. Local patterns of regression coefficients from the GLM on unsmoothed data were extracted from each region for each of the 15 mental states in the study. These local activity patterns were vectorized and correlated to measure their similarity. The lower triangular elements of the resulting correlation matrix thus represented the neural similarity between each of the 15 states. A separate set of participants provided ratings for how likely each of the 15 states is to transition to every other state (Thornton and Tamir, 2017). The neural similarities were vectorized and correlated with the transitional probability ratings. This correlation measured the extent to which neural patterns for an individual state resembled the states to which it is likely transition. Because the pattern similarity measure (correlation) was necessarily symmetric, we averaged transitional probabilities across the diagonal before correlating these ratings with pattern similarity. This procedure was repeated with the searchlight centered at every voxel in the image provided it contained at least 30 voxels. The result was a whole-brain map of correlation coefficients for each participant. These correlation maps were smoothed with a 6 mm FWHM Gaussian kernel and then entered into a one-sample *t* test (vs zero) to assess statistical significance at the group level. The results were corrected for multiple comparisons ($p < 0.05$) using maximal statistical permutation testing with threshold free cluster enhancement (TFCE; Smith and Nichols, 2009).

We conducted the same representational similarity analyses within the social brain network as a whole. This network was defined independently using a mask consisting of 10,216 (noncontiguous) voxels implicated in social cognition. Previous research indicated that these voxels maximized both voxelwise and patternwise reliability of neural activity when mentalizing about a large group of people (Thornton and Mitchell, 2018). Within these voxels, we repeated the analysis performed at the searchlight level to measure the extent to which neural similarity across the entire network reflected transitional probabilities. We tested both average transition ratings from an independent set of participants (Thornton and Tamir, 2017), as well as the ratings of individual participants in the imaging study. In both cases, we calculated statistical significance by Fisher transforming the resulting correlations and performing a one-sample *t* test (vs zero) at the group level.

Asymmetric representational similarity analysis. Typical representational similarity analyses such as those in the previous section use symmetric metrics such as a correlation matrix to estimate the similarity between patterns of brain activity. This approach suffices when theoretical predictions are likewise symmetric. However, in the present case, transitional probabilities are meaningfully asymmetric (e.g., "drunkenness" is more likely to precede "sleepiness" than to follow it). To further test our hypothesis that patterns of brain activity reflexively encode the transitional probabilities between mental states, we developed a novel extension of representational similarity analysis that relies on an asymmetric similarity measure from information theory: Kullback–Leibler (KL) divergence, also known as relative entropy.

KL divergence is typically used to compare different probability distributions to each other. It can be thought of as a measure of how much information is gained when updating from one distribution to another. It is the directed nature of this comparison that gives KL divergences its potential for asymmetry. So, for example, a normal distribution is more similar to (i.e., less divergent from) a uniform distribution than a uniform distribution is to a normal distribution. The reason is that uniform

distribution embodies very little information (i.e., a flat prior, a representation of complete ignorance, in Bayesian terms), so relatively more information is gained when updating from the uniform distribution to the binomial distribution. In contrast, updating from the normal distribution to the uniform distribution involves minimal information gain and thus a smaller KL divergence (i.e., higher similarity).

We computed neural pattern similarity by calculating KL divergence between the patterns corresponding to each pair of mental states. In preparation for this procedure, neural patterns were transformed to resemble the expected input to the empirical KL divergence function in the "entropy" package in R. Specifically, all values in each pattern were made positive by subtracting the minimum for each pattern. The decimals were then converted to integers without loss of precision by multiplying each pattern by the minimum absolute difference between any pair of values within a given pattern for each participant and then rounding. Finally, a constant value of 1 was added to each pattern to avoid division by 0.

The similarity values resulting from KL divergence consist of both symmetric and asymmetric components. To isolate the asymmetric components for further analysis, we first computed the purely symmetric component by averaging the asymmetric neural pattern similarity matrix with its transpose. We then subtracted this symmetric component from the original partially asymmetric similarity matrix to produce a purely asymmetric matrix that was orthogonal to the symmetric component. We applied the same procedure to the transitional probability matrices to likewise compute their asymmetric components. This included both the transitional probabilities provided by independent raters and the transitional probabilities provided by the imaging participants in the present study.

Finally, we correlated the transitional probability asymmetries with the neural pattern similarity asymmetries, leaving out the diagonal of each matrix. Due to the computationally intensive nature of the computing the KL divergence, we performed this asymmetric pattern analysis only once, at the level of the social brain network as a whole, and not within searchlights throughout the brain. Group level significance testing was performed via *t* tests on Fisherized correlation coefficients, as in the traditional symmetric representational similarity analysis.

Controlling for perceived similarity. The traditional and asymmetric representational similarity analyses described above test the association between transitional probability ratings and neural pattern similarity. However, in previous research, we found that transitional probability ratings were highly related with perceptions of similarity between mental states (Thornton and Tamir, 2017). We suggested in that work that this is likely because people form their intuitions about state similarity based, at least in part, on mental state dynamics. That is, people may judge two states to be similar because they regularly co-occur or follow one another. If so, then statistically controlling for perceived similarity when measuring the association between transitional probability and neural pattern similarity would produce misleading results. However, the nature of causation between state transitions and perceived similarity has not yet been empirically established. To the extent that people make their transitional probability judgements based on perceived similarity, rather than vice versa, perceived similarity constitutes a potential confound in the present analyses.

To address this issue, we repeated the social brain network representational similarity analyses while statistically controlling for perceived similarity. Perceived similarity ratings were provided by a separate group of participants and these values were averaged across participants to provide a single set of estimates of the perceived similarities between states. We then repeated the symmetric and asymmetric representational similarity analyses described in the preceding sections, controlling for any association between perceived similarity and neural pattern similarity (i.e., regressing out the influence of perceived similarity on neural pattern similarity, and correlating the residuals of these regressions with transitional probability ratings).

The symmetric version of this analysis cannot distinguish between the two accounts of perceived similarity described above (i.e., similarity as either the effect, or as the cause, of transitional probabilities). Both of these accounts imply a strong association between perceived similarity and transitional probability ratings, and (potentially full) statistical mediation of the latter by the former. However, the same cannot be said of the asymmetric representational similarity analysis. Asymmetries in transitional probabilities represent regularities in mental state dynamics, whereas asymmetries in the perceived similarity between concepts are thought to result from the concepts in question having different numbers of features associated with them (Tversky, 1977). Because mental state dynamics have no prima facie connection with the number of features possessed by each state, we expected the asymmetric components of transitional probability ratings and perceived similarity ratings to diverge. This expectation made the asymmetric representational similarity analysis the crucial test of whether the present analyses should be interpreted in terms of predictive coding or the perceived similarity between mental states.

State frequency representational similarity analysis. Knowledge of the transitional probabilities between different mental states would be a major boon to social prediction. The long-run frequencies of mental states are a downstream consequence of those transitional probabilities. For instance, if a person highly likely to transition into happiness, they will spend more time in this state than a person unlikely to transition into happiness. The expected frequencies that result from transitional probabilities are useful in their own right for making social predictions. Therefore, if the brain tracks the transitional probabilities between mental states, then one would expect it to also encode the resulting long-term state frequencies.

We tested this hypothesis by computing expected state frequencies from the average transitional probability ratings provided by independent raters (Thornton and Tamir, 2017). These expected state frequencies, known as the stationary distribution of the Markov chain, can be computed by raising a transitional probability matrix to a high exponential power (i.e., in this case, multiplying it by itself 100 times). Before this procedure, the rows of the rated transitional probability matrix were normalized to sum to one, to create a valid Markov chain. The result of the exponentiation procedure was a matrix with 15 identical rows, with each row representing the expected state frequencies of the 15 mental states. These values were converted to a similarity matrix by taking the reverse-coded absolute differences between each pair of state frequencies. The resulting similarity estimates were entered into a representational similarity analysis in which they were correlated with the symmetric component of neural pattern similarity in the social brain network. A positive correlation would indicate that mental states with similar frequencies elicit similar patterns of brain activity or, more simply, that the social brain automatically encodes the expected frequencies of mental states. Given that the frequency estimates were derived from transitional probability ratings, we also repeated this representational similarity analysis while controlling for the association between neural pattern similarity and transitional probabilities to rule out the possibility that apparent frequency effects were merely due to the transitions themselves.

Note that the procedure for calculating expected state frequencies is only well defined if the initial matrix in question is a transitional probability matrix. Therefore, if transitional probability ratings are merely a proxy for perceived similarity, then we would expect this representational similarity analysis to show no effect. It is possible that raising a similarity matrix to a high exponential power and then extracting its first row might have some meaningful interpretation but, if so, to our knowledge, it has yet to be described. Assuming that it does not, then we would not expect the brain to encode this information unless similarity is shaped by transitional probability rather than vice versa. Therefore, in addition to testing a different facet of predictive coding, this analysis can also indirectly help to arbitrate between a transitional probability interpretation and a perceived similarity interpretation of the present set of findings.

Repetition suppression. The neuroimaging data were also analyzed to determine whether voxelwise activity was consistent with repetition suppression in response to expected mental state sequences. Repetition suppression provides a valuable addition to the pattern analyses for two reasons. First, it relies on a different form of signal, univariate activity, instead of multivariate patterns and thus provides a partially independent test of the predictive coding hypothesis. Second, like asymmetric

representational similarity analysis, repetition suppression can capture asymmetric transitions between states, whereas traditional pattern analyses cannot.

If the social brain uses current states to reflexively predict future states, then expected sequences (e.g., "drunkenness" followed by "sleepiness") should elicit repetition suppression, whereas unexpected sequences (e.g., "sleepiness" followed by "drunkenness") should not. To test this hypothesis, we repeated the GLM described above (i.e., with 15 boxcar regressors convolved with a canonical HRF to model the 15 mental states in the study). Smoothed preprocessed data (6 mm FWHM) were used because fine-grained spatial resolution was not necessary for this analysis. In addition to the condition regressors, this GLM also contained 15 parametric modulators, one for each mental state, to model the effects of repetition suppression. The values of the modulators were fixed to the values of the transitional probability from the state on the previous trial to the state on the current trial. Following the GLM, the regression coefficient maps from the 15 parametric modulators were averaged and the resulting averages were entered into voxelwise one-sample *t* tests (vs zero) across participants. This analysis targeted the negative direction of the test because higher transitional probabilities were predicted to produce less activity. As in the searchlight analysis, the results were corrected for multiple comparisons using maximal statistical permutation testing with TFCE (Smith and Nichols, 2009).

Dimensional mediation. Finally, we tested whether a simple set of psychological dimensions could explain the extent to which transitional probabilities predict neural pattern similarity (Tamir et al., 2016; Thornton and Tamir, 2017; Tamir and Thornton, 2018). That is, we tested whether proximity on four psychological dimensions statistically mediated the relation between pattern similarity and transitional probability ratings assessed above. If so, then this would suggest that these dimensions provide a natural scaffolding for social prediction: merely encoding a state's position on these dimensions would imply its likely transitions.

We performed this mediation analysis in three steps. In step 1, we examined whether these psychological dimensions could describe neural pattern similarity. To do so, we computed proximity measures (reverse-coded absolute differences) between each pair of states using ratings of four dimensions: rationality (cognitive vs emotional states), social impact (socially arousing vs nonsocial/low-arousal states), valence (positive vs negative states), and human mind (uniquely human, purely mental states vs those shared with other animals or with a somatic component). We regressed neural pattern similarity onto these dimension-derived proximity measures. This regression allowed us to determine the extent to which each dimension could uniquely explain the similarity between neural representations of mental states. As in the primary representational similarity analysis, we computed this regression separately for each imaging participant and then tested whether these results were significant via one-sample *t* tests across the regression coefficients. Here, a significant association between neural pattern similarity and distance on each dimension except human mind would replicate prior research demonstrating the role of these dimensions in mental state representation (Tamir et al., 2016).

In step 2, we measured the residual relationship between neural pattern similarity and transitional probability after controlling for proximity on these psychological dimensions. If the dimensions statistically mediate the relationship between pattern similarity and transitional probability, then this residual relationship should be significantly smaller than the full (zero-order) correlation. To measure this residual relationship, we correlated the transitional probabilities ratings with residual pattern similarity; that is, pattern similarity after the effect of the dimensions was removed by the regression from step 1. This produced a set of semipartial correlations (one per imaging participant) between residual neural pattern similarity (with dimension-related variance removed) and transitional probability. These semipartial correlations reflect the strength of the association between pattern similarity and transitional probability when the role of the dimensions is fully accounted for.

In step 3, we completed the analysis by testing whether the association between neural pattern similarity and transitional probability became significantly smaller when the variance associated with the psychological dimensions was removed from the neural pattern similarity. If so, then this would indicate that part of the relationship between pattern similarity and transitional probability could potentially be attributed to proximity on these dimensions. That is, by encoding current states using these dimensions, the brain might simultaneously encode likely future states because future states are located nearby on those dimensions. To test this hypothesis, we compared the zero-order correlation between pattern similarity and transitional probability to the semipartial correlation. This difference (Δr) was calculated independently for each participant and the significance of the overall difference was tested using one-sample *t* tests on Fisher-transformed correlations. Here, a significant result indicates that proximity in the dimensional space explains the shared variance between pattern similarity and transitional probability.

Finally, in addition to testing whether the four psychological dimensions statistically mediate the relation between pattern similarity and transitional probabilities, we also tested whether this mediation was complete or only partial. To do so, we calculated the significance of the residual relationship from step 2 via one-sample *t* test on Fisher-transformed semipartial correlation coefficients. If the residual relationship was significant, then this would indicate that the mediation is partial rather than complete. That is, a significant residual relationship would indicate that the relationship between neural pattern similarity and transitional probability cannot be completely explained by proximity on the psychological dimensions that we consider.

Results

Behavioral results

Behavioral data were analyzed to assess the quality of participants' responses during the imaging experiment. The average response rate in the imaging task was 95% (SD = 0.07), indicating high participant engagement. The average response was 4.26 (SD = 0.37) of 5, indicating that the scenarios succeeded in representing their respective mental states. The average SD of responses with respect to states was only 0.27 (SD = 0.12), indicating that the genetic algorithm was successful in choosing scenarios that were equally appropriate for each state. Participants' first and second ratings of the same scenario were consistent (mean $r = 0.39$, $d = 1.93$, $p = 9.5 \times 10^{-11}$). The average response time was 1.47 s (SD = 0.29 s) from when the scenario appeared on each trial. Reported SDs were calculated at the participant level. Together, the behavioral results demonstrate that participants were consistently and sensibly engaging with the fMRI task.

Representational similarity analysis

Representational similarity analysis was used to test whether activity patterns for each current mental state resembled patterns for likely future states more than patterns for unlikely future states. Searchlight representational similarity analysis revealed a single brain region in the dorsal medial prefrontal cortex that manifested neural pattern similarity consistent with the transitional probabilities between mental states (Fig. 1*A*). This region had an extent of 666 voxels at a permutation TFCE-corrected threshold of $p < 0.05$, with a peak voxel ($p_{corrected} = 0.011$) at $x = -6$, $y = 48$, $z = 14$ in MNI coordinates.

The same representational similarity analysis then was repeated within an independently defined social brain network of 10,216 voxels (Fig. 2*A*; Thornton and Mitchell, 2018). The selected regions closely resembled those typically implicated in social cognition (Van Overwalle and Baetens, 2009), including medial prefrontal and parietal cortices, the anterior temporal lobe, and the temporoparietal junction. As expected, neural patterns within these regions reflected transitional probabilities, such that states with higher transitional probabilities between them elicited more similar patterns (Fig. 2*B*). This relationship was robust when using transitional probability ratings from

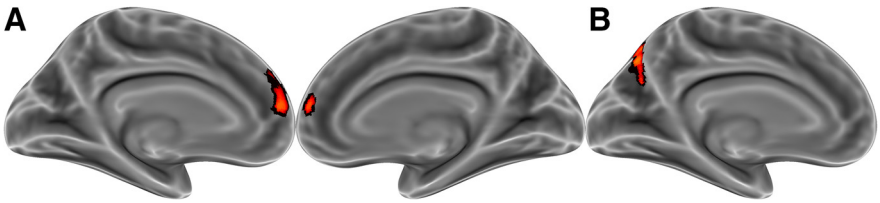

Figure 1. Whole-brain mapping of the neural representation of transitional probabilities between states. ***A***, Transitional probability judgments correlated with neural pattern similarity in medial prefrontal cortex. ***B***, The same transitional probability judgments predict repetition suppression in the posterior precuneus. Results are corrected for multiple comparisons ($p < 0.05$).

an independent sample of participants (mean $r = 0.13$, $d = 0.95$, $p = 0.00003$) or when using transition ratings provided by participants in the imaging experiment (mean $r = 0.13$, $d = 1.07$, $p = 0.000006$). Together, these findings demonstrate that, whenever a person considers a given mental state, this evokes a neural pattern that resembles the patterns of likely future states. That is, future states are automatically encoded in representations of current states.

Asymmetric representational similarity analysis

Asymmetric representational similarity analysis revealed a significant relationship between asymmetries in transitional probability ratings and asymmetries in neural pattern similarity within the social brain network (mean $r = 0.10$, $d = 0.78$, $p = 0.0003$). This effect was also observed when analyzing transitional probability ratings made by individual imaging participants instead of independent raters (mean $r = 0.07$, $d = 0.54$, $p = 0.008$). These results indicate that asymmetries in neural pattern similarity encode asymmetric mental state dynamics (i.e., cases in which one state is more likely to follow another state than to precede it). The sizes of these effects are similar to those observed in the symmetric representational similarity analyses reported in the previous section, suggesting that the symmetric and asymmetric components of neural pattern similarity are approximately equally engaged in representing others' future states.

Controlling for perceived similarity

Symmetric and asymmetric representational similarity analyses were repeated while statistically controlling for ratings of the perceived similarity between mental states. We hypothesized that: (a) the social brain automatically predicts others' future states, (b) transitional probabilities play a causal role in forming impression of perceived similarity between mental states, and (c) asymmetries in transitional probability and perceived similarity results from different sources (asymmetric mental state dynamics and different numbers of features associated with each state, respectively). If these three hypotheses are true, then we would expect full statistical mediation of the transitional probabilities by perceived similarity in the symmetric analysis (due to hypothesis b), but no mediation of transitional probability by perceived similarity in the asymmetric analysis (due to hypothesis c).

This is precisely the pattern of results that we observed: in the symmetric analysis, transitional probabilities were no longer a significant predictor of neural pattern similarity when controlling perceived similarity (mean $r = -0.004$, $d = -0.10$, $p = 0.58$). However, in the asymmetric analysis, transitional probabilities remained a significant predictor of neural pattern similarity (mean $r = 0.10$, $d = 0.78$, $p = 0.0003$). Again, we observed similar results when using transition ratings from each imaging participant instead of independent raters for both the symmetric analysis (mean $r = 0.03$, $d = 0.35$, $p = 0.07$) and asymmetric analysis (mean $r = 0.07$, $d = 0.54$, $p = 0.008$).

The magnitudes of the asymmetric effects were almost identical to those in the reported in the previous section, indicating similarity ratings accounted for virtually none of the shared variance between asymmetric neural pattern similarity and asymmetric transitional probability ratings. Ratings of the perceived similarity between mental states were highly reliable ($\alpha = 0.90$), indicating that these results reflect strong statistical control, rather than measurement error in the covariate (Westfall and Yarkoni, 2016). Together, these results indicate that the present results cannot be attributed to confounding transitional probability with perceived similarity. In particular, the critical test case of asymmetric neural pattern similarity demonstrates the incremental validity of transitional probabilities over and above perceived similarity.

State frequency representational similarity analysis

Consistent with the hypothesis that the social brain encodes information useful for predicting others' states, we found that (symmetric) neural pattern similarity was significantly associated with similarity in the expected frequencies of mental states (mean $r = 0.10$, $d = 1.18$, $p = 0.000001$). That is, states with similar frequencies were encoded with similar neural patterns. Additionally, this finding replicated when statistically controlling for transitional probabilities (mean $r = 0.09$, $d = 1.10$, $p = 0.000003$). That is, even though the expected frequencies in this analysis were derived from the stationary distribution of the rated transitional probabilities, these frequency expectations independently predict neural pattern similarity. This result also indirectly supports the interpretation of the present findings in terms of social prediction rather than perceived similarity because applying the procedure to calculate a stationary distribution to an ordinary similarity matrix would produce results with no clear meaning and which one would not expect to correlate with neural pattern similarity.

Repetition suppression

Repetition suppression analysis was used to provide convergent evidence for the hypothesis that the engages in reflexive mental state prediction via a process similar to predictive coding. If the brain automatically predicts others' future states, then the brain should require less processing when states are shown in predictable (vs unpredictable) sequences. That is, predicted states should elicit less neural activity than unpredicted states. As hypothesized, predictable sequences of states elicited less activity in a region within the social brain network (Fig. 1*B*). This region in the posterior precuneus had an extent of 298 voxels at a permutation TFCE-corrected threshold of $p < 0.05$, with a peak voxel ($p_{\text{corrected}} = 0.008$) at $x = -10$, $y = -74$, $z = 40$ in MNI coordinates. This finding further reflects the centrality of prediction to social cognition: the social brain automatically encodes predictions of others' future states.

Dimensional mediation

Thus far, our results demonstrate that the brain makes automatic predictions about others' future states. Next, we tested hypotheses derived from our theoretical model of the predictive social mind for how people make these predictions (Tamir and Thorn-

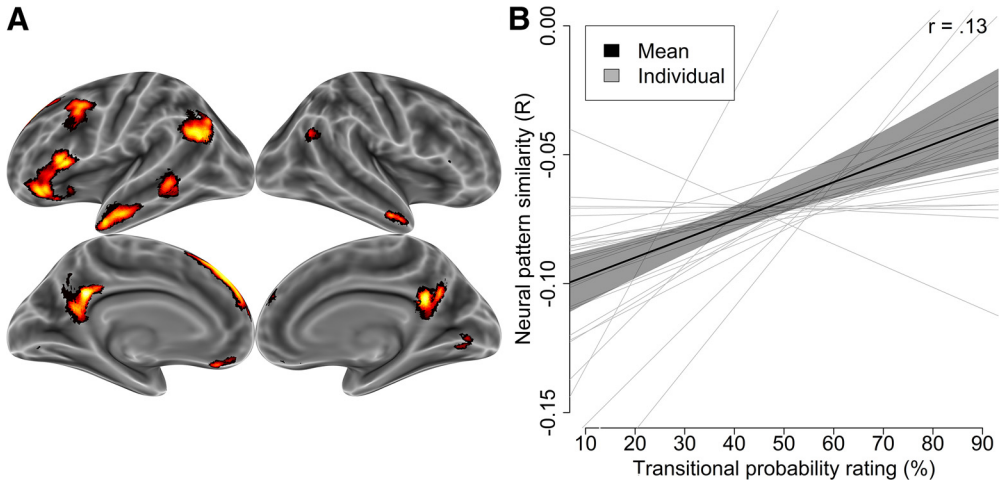

Figure 2. Transitional probabilities were associated with neural pattern similarity in the social brain network. ***A***, Social brain network analyses were conducted within these independently defined regions. ***B***, Neural patterns for each state resembled the pattern of states to which it was likely to transition. Transitional probabilities between mental states correlated with neural pattern similarity between those states. Each participant was analyzed as an independent unit (light gray lines) and then averaged (black line). The shaded region around the mean slope represents the bootstrapped 95% confidence interval.

ton, 2018). Specifically, we tested whether three psychological dimensions, rationality, social impact, and valence, might scaffold social predictions. To do so, we first replicated earlier work showing that the brain represents mental states using these three psychological dimensions (Tamir et al., 2016). Distance on each dimension correlated with neural pattern similarity between mental states (rationality: mean $\beta = 0.04$, $d = 0.67$, $p = 0.001$; social impact: mean $\beta = 0.16$, $d = 1.14$, $p = 0.000002$; and valence: mean $\beta = 0.07$, $d = 0.71$, $p = 0.0009$). A fourth psychological dimension, the human mind, was again not associated with pattern similarity (mean $\beta = 0.02$, $d = 0.18$, $p = 0.34$).

Next, we tested whether the brain uses these dimensions, not only to represent static mental states, but also to make mental state predictions. The correlation between transitional probability and neural pattern similarity was significantly reduced after regressing out the influence of these four dimensions (mean $\Delta r = 0.09$, $d = 1.06$, $p = 0.000007$). This result suggests that the brain makes predictions of others' futures states by using the positions of states on the dimensions of rationality, social impact, and valence. We then tested whether this statistical mediation was full or partial. The residual relationship remained significant (mean $r = 0.03$, $d = 0.52$, $p = 0.01$), suggesting that the dimensions do not fully mediate the relation between neural pattern similarity and transitional probability. These results replicated those using transitional probability ratings from individual participants in the imaging experiment: the relationship between transitional probability and pattern similarity was mediated by the psychological dimensions (mean $\Delta r = 0.07$, $d = 0.98$, $p = 0.00002$) and the residual relationship remained significant (mean $r = 0.06$, $d = 0.77$, $p = 0.0004$). Together, these findings support the theoretical model, suggesting that people represent others' minds using a low-dimensional representational space and that they could leverage this dimensional representation to predict others' social futures.

Discussion

The current findings provide convergent evidence for the hypothesis that glimpses of the social future are directly incorporated into representations of the here and now. Whenever we think about another person in a particular mental state, our brains automatically generate a prediction of their social future. These predictions are built into multivariate representations of others' current states, such that thinking about a given mental state elicits a neural pattern that literally resembles the neural pattern for states that typically follow it. Likewise, viewing mental states in a predictable order results in univariate repetition suppression. These findings underscore the centrality of prediction to social cognition and mental state representation (Koster-Hale and Saxe, 2013; Barrett, 2017) by demonstrating that the brain reflexively predicts others' future mental states.

The current findings support the theory that the mind leverages a 3D structure of mental state representation to predict others' mental states (Tamir and Thornton, 2018). The current findings replicate earlier work demonstrating that people represent others' minds using a low-dimensional space such that proximity between states on three psychological dimensions, rationality, social impact, and valence, correlates with neural pattern similarity (Tamir et al., 2016). Other research demonstrates that proximity on these same dimensions predicts actual and perceived transitional probabilities between states (Thornton and Tamir, 2017). Here, we connect these findings by demonstrating that proximity between mental states mediates the association between transitional probabilities and neural pattern similarity. That is, when people's brains encode states using these dimensions, they thereby reflexively and efficiently predict others' future mental states. These results suggest that, when we perceive other people, we represent not only the static features of their current state, but also automatically predict their social futures, all using the same parsimonious dimensional space. However, statistical mediation by the psychological dimensions is ultimately correlational. Therefore, although the observed statistical mediation is consistent with our model of predictive social cognition, this aspect of the model still awaits definitive support from future experiments.

The current study design allows for a strong inference that the brain automatically predicts others' mental states. Participants were never instructed to make social predictions during the neuroimaging task. Participants were shown states one at a time, in a seemingly arbitrary order; they rated the likelihood of state transitions only after completing the imaging task. This task neither required nor rewarded knowledge of transitional probabilities, so

participants had no extrinsic reason to attend to these probabilities. Therefore, the observation that neural representations of current mental states resemble those of future states supports the claim that people generate these predictions spontaneously. That said, the current transition rating task and fMRI paradigm are highly controlled and not necessarily representative of naturalistic social prediction. Although this task offers strong proof-of-concept that our brain reflexively predicts others' future states, it will be important to replicate these findings using more realistic paradigms. Future research that situates state-to-state prediction within the broader context of everyday social cognition will offer more generalizable conclusions about social prediction (Saxe, 2018).

Both whole-brain representational similarity and repetition suppression analyses implicated portions of the social brain network in representing state transitions; however, the specific regions implicated differ: dorsal medial prefrontal cortex and the posterior precuneus, respectively. Both of these regions are part of the putative social brain network, a set of regions that is reliably engaged when people engage in social thought (Van Overwalle and Baetens, 2009). Recent research has implicated this network as a whole in encoding social prediction errors when reasoning about others' minds (Theriault et al., 2017). However, in that study, predictive coding effects did not vary significantly across different portions of the network. In contrast, the present results suggest a degree of spatial specify to the predictive coding functions. That said, whereas the present findings bear many hallmarks of predictive coding, it is important to note that fMRI lacks the spatial specificity to test whether predictive coding per se, or merely something like it, is implemented at level of individual neurons (Barrett and Simmons, 2015).

This returns us to the question of why representational similarity analysis and repetition suppression implicate different regions in social cognition. Research on visual working memory suggests that pattern analysis may track explicit representations, whereas repetition suppression reflects implicit representations (Ward et al., 2013). If this finding generalizes to social cognition, it would suggest that the precuneus might respond more robustly to implicit measures of perceived transitional probabilities, which might lack some of the response biases that can contaminate explicit measures. A more substantive explanation might involve differential functioning. For instance, it is possible that medial prefrontal cortex is responsible for maintaining mental models of others' minds, but that the precuneus is responsible for comparing the predictions made by this model with observed sequences of real-world events. This putative function of medial prefrontal cortex would be broadly consistent with other work implicating this region in representing other predictive mental models, such as schema for events (Baldassano et al., 2018) and prospective memory for future tasks (Momennejad and Haynes, 2013). Notably absent from the current findings are regions habitually implicated in nonsocial prediction error, such as the basal ganglia, anterior cingulate, and ventromedial prefrontal cortex (Garrison et al., 2013). This discrepancy might be explained by differences in content (social vs nonsocial), the absence of explicit reward in the present study, or analytic differences (pattern analysis/repetition suppression vs univariate analyses).

The novel asymmetric extension of representational similarity analysis further supports the automatic social prediction hypothesis. This analysis demonstrated that neural patterns encode asymmetries in the transitional probabilities between mental states. Moreover, this relationship remained significant and nearly unchanged in magnitude when statistically controlling for the perceived similarity between mental states. This incremental validity of transitional probabilities over and above perceived similarity helps to rule out the possibility that the latter could explain away the present results. We have previously explored the extent to which state transitions can be predicted by state similarity (Thornton and Tamir, 2017): transitional probability and similarity judgments are closely related, but the latter cannot fully explain the former. If there is a direct causal relationship between transitional probability and holistic similarity, then we propose that the former may well cause the latter rather than the reverse. That is, people may judge the similarity between two states based on the likelihood that one co-occurs with or follows another. We hope that future work will test the extent to which perceptions of transition likelihood drive perceptions of similarity rather than vice versa.

The present results also suggest that the social brain encodes the expected frequencies of mental states: states with similar frequencies were found to elicit similar neural activity patterns. Like transitional probabilities, the expected frequencies or base rates of mental states could be highly useful for making accurate social predictions. These frequencies can themselves be thought of as a consequence of the transitional probability matrix: if a person is highly likely to transition into state A and unlikely to transition into state B, then we would expect them to experience state A more frequently than state B. This is in fact how we derived expectations of state frequencies: by computing the stationary distribution of the Markov chain comprised by transitional probability ratings. These frequencies expectations remain a significant predictor of neural similarity even when controlling for transitional probability, indicating that they are independently encoded by patterns of neural activity in the social brain network. Moreover, because the process that we used to derive expected frequencies is only well defined if the underlying matrix represents transitional probabilities rather than similarity more generally, this finding indirectly supports the interpreting the present results in terms of social prediction rather than perceived similarity.

The present investigation considers only one-step transitions and the long-run frequencies derived from them. However, future work might productively investigation Markov chains with "memory"; that is, transitional probability matrices in which the two preceding states influence the likelihood of the next state. Likewise, the neural representation of mental state co-occurrence remains unexplored. Like one-step and "memory" Markov chains, co-occurrences can be viewed as a special case of transitional probability: zero temporal lag. Previously, we found a large association between one-step transitional probabilities and co-occurrences, suggesting that the present results may represent a combination of these different lags (Thornton and Tamir, 2017). However, because co-occurrences are necessarily symmetric, they cannot explain the asymmetric representational similarity results. In addition to considering different types of transitional probabilities, future work might also investigate how these transitional probabilities are processed and cached for efficient social prediction, for instance, via successor representations (Momennejad et al., 2017).

The current findings inform both neural and psychological perspectives on social cognition. From a neural point of view, this outcome suggests that a shared principle, the goal of prediction, describes not only sensory cortices, but also those brain regions implementing the most abstract functions such as understanding other people's mental states. From a psychological point of view,

the finding that psychological dimensions scaffold social prediction provides an additional raison d'être for many data-driven theories of social content. Theories of this sort, such as the circumplex model of affect (Russell, 1980; Posner et al., 2005), have long proven useful for making sense of their respective domains. The current findings suggest that the dimensions supporting static theories might also facilitate dynamic prediction. Focusing on prediction may thus motivate and unify existing social psychological theories. In sum, social predictive coding offers an integrative and generative framework for understanding the organization of social knowledge and how people draw upon it to glimpse the social future (Tamir and Thornton, 2018).

References

Aguirre GK (2007) Continuous carry-over designs for fMRI. Neuroimage 35:1480–1494. CrossRef Medline

Aguirre GK, Mattar MG, Magis-Weinberg L (2011) de Bruijn cycles for neural decoding. Neuroimage 56:1293–1300. CrossRef Medline

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113. CrossRef Medline

Baldassano C, Hasson U, Norman KA (2018) Representation of real-world event schemas during narrative perception. J Neurosci 38:9689–9699. CrossRef Medline

Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I (2001a) The "Reading the mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. J Child Psychol Psychiatry 42:241–251. CrossRef Medline

Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E (2001b) The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. J Autism Dev Disord 31:5–17. CrossRef Medline

Barrett LF (2017) The theory of constructed emotion: an active inference account of interoception and categorization. Soc Cogn Affect Neurosci 12:1–23. CrossRef Medline

Barrett LF, Simmons WK (2015) Interoceptive predictions in the brain. Nat Rev Neurosci 16:419–429. CrossRef Medline

Barrett LF, Mesquita B, Gendron M (2011) Context in emotion perception. Current Directions in Psychological Science 20:286–290. CrossRef

Becchio C, Koul A, Ansuini C, Bertone C, Cavallo A (2018) Seeing mental states: An experimental strategy for measuring the observability of other minds. Phys Life Rev 24:67–80. CrossRef Medline

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 36:181–204. CrossRef Medline

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci 364:1211–1221. CrossRef Medline

Garrison J, Erdeniz B, Done J (2013) Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 37:1297–1310. CrossRef Medline

Gregory RL (1980) Perceptions as hypotheses. Philos Trans R Soc Lond B Biol Sci 290:181–197. CrossRef Medline

Hohwy J, Roepstorff A, Friston K (2008) Predictive coding explains binocular rivalry: an epistemological review. Cognition 108:687–701. CrossRef Medline

Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM (2012) FSL. Neuroimage 62:782–790. CrossRef Medline

Kilner JM, Friston KJ, Frith CD (2007) Predictive coding: an account of the mirror neuron system. Cogn Process 8:159–166. CrossRef Medline

Koster-Hale J, Saxe R (2013) Theory of mind: a neural prediction problem. Neuron 79:836–848. CrossRef Medline

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868. CrossRef Medline

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis–connecting the branches of systems neuroscience. Front Syst Neurosci 2:4. CrossRef Medline

Litman L, Robinson J, Abberbock T (2017) TurkPrime.com: a versatile crowdsourcing data acquisition platform for the behavioral sciences. Behav Res Methods 49:433–442. CrossRef Medline

Momennejad I, Haynes JD (2013) Encoding of prospective tasks in the human prefrontal cortex under varying task loads. J Neurosci 33:17342–17349. CrossRef Medline

Momennejad I, Russek EM, Cheong JH, Botvinick MM, Daw ND, Gershman SJ (2017) The successor representation in human reinforcement learning. Nature Human Behaviour 1:680–692. CrossRef

Posner J, Russell JA, Peterson BS (2005) The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 17:715–734. CrossRef Medline

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87. CrossRef Medline

Raskin RN, Hall CS (1979) A narcissistic personality inventory. Psychology Reports 45:590. CrossRef Medline

Russell DW (1996) UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 66:20–40. CrossRef Medline

Russell JA (1980) A circumplex model of affect. Journal of Personality and Social Psychology 39:1161–1178. CrossRef

Saxe R (2018) Seeing other minds in 3D. Trends Cogn Sci 22:193–195. CrossRef Medline

Sherbourne CD, Stewart AL (1991) The MOS social support survey. Soc Sci Med 32:705–714. CrossRef Medline

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83–98. CrossRef Medline

Tamir DI, Thornton MA (2018) Modeling the predictive social mind. Trends Cogn Sci 22:201–212. CrossRef Medline

Tamir DI, Thornton MA, Contreras JM, Mitchell JP (2016) Neural evidence that three dimensions organize mental state representation: rationality, social impact, and valence. Proc Natl Acad Sci U S A 113:194–199. CrossRef Medline

Theriault J, Young L, Theriault J (2017) Social prediction in the theory of mind network. Available at https://psyarxiv.com/hvn54/.

Thornton MA, Mitchell JP (2018) Theories of person perception predict patterns of neural activity during mentalizing. Cereb Cortex 28:3505–3520. CrossRef Medline

Thornton MA, Tamir DI (2017) Mental models accurately predict emotion transitions. Proc Natl Acad Sci U S A 114:5982–5987. CrossRef Medline

Tversky A (1977) Features of similarity. Psychol Rev 84:327–352. CrossRef

Van Overwalle F, Baetens K (2009) Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage 48:564–584. CrossRef Medline

von Helmholtz H (1910) Treatise on physiological optics. North Chelmsford, MA: Courier Corporation.

Vuust P, Ostergaard L, Pallesen KJ, Bailey C, Roepstorff A (2009) Predictive coding of music–brain responses to rhythmic incongruity. Cortex 45:80–92. CrossRef Medline

Ward EJ, Chun MM, Kuhl BA (2013) Repetition suppression and multivoxel pattern similarity differentially track implicit and explicit visual memory. J Neurosci 33:14749–14757. CrossRef Medline

Westfall J, Yarkoni T (2016) Statistically controlling for confounding constructs is harder than you think. PLoS One 11:e0152719. CrossRef Medline

Zaki J, Bolger N, Ochsner K (2009) Unpacking the informational bases of empathic accuracy. Emotion 9:478–487. CrossRef Medline